Image Retrieval by Content: A Machine Learning Approach

Usama M. Fayyad Padhraic Smyth

Jet Propulsion Laboratory, M/S 525-3660
California Institute of Technology
Pasadena, CA 91109
{fayyad,pjs}@aig.jpl.nasa.gov

ABSTRACT

In areas as diverse as Earth remote sensing, astronomy, and medical imaging, there has been an explosive growth in the amount of image data available for creating digital image libraries. However, the lack of automated analysis and useful retrieval methods stand in the way of creating true digital image libraries. In order to perform query-by-content type searches, the query formulation problem needs to be addressed: it is often not possible for users to formulate the targets of their searches in terms of queries. We present a natural and powerful approach to this problem to assist scientists in exploring large digital image libraries. We target a system that the user trains to find certain patterns by providing it with examples. The learning algorithms use the training data to produce classifiers to detect and identify other targets in the large image collection. This forms the basis for query by content capabilities and for library indexing purposes. We ground the discussion by presenting two such applications at JPL: the SKICAT system used for the reduction and analysis of a 3 terabyte astronomical data set, and the JARtool system to be used in automatically analyzing the Magellan data set consisting of over 30,000 images of the surface of Venus. General issues which impact the application of learning algorithms to image analysis applications are discussed.

Keywords: machine learning, pattern recognition, automated data analysis and archival, large image databases, query by content.

1 INTRODUCTION

Image acquisition technology has undergone tremendous improvements in recent years. The vast amounts of scientific data, stored in the form of digital image libraries, are potential treasure-troves for scientific investigation and analysis. Unfortunately, advances in our ability to deal with this volume of data in an effective manner have not paralleled the hardware gains in storage technology and data gathering instruments. While special-purpose tools for particular applications exist, no general-purpose software tools and algorithms which can assist a scientist in exploring large scientific image libraries are available. This paper presents our recent progress in developing interactive semi-automated image library exploration tools based on pattern recognition and machine learning techniques. Two successful applications at JPL will be used to ground the discussion and point out the powerful impact such learning tools can have. We then proceed to discuss the general problem of automated image library exploration, the particular aspects of image databases which distinguish them from other databases, and how this impacts the application of off-the-shelf learning algorithms to problems of this nature.

We are developing tools that can be trained by example to execute difficult query-by-content type tasks on large image databases. A scientist provides training examples by locating candidate targets within an image on the screen. The learning algorithms use the training data to produce a classifier that will detect and identify other targets in a large library of similar images. The scientist can thus customize the tool to search for one type of visual feature versus another simply by providing positive and negative examples. This is a non-intrusive method in the sense that it will not require the scientists to perform anything different from what they do now; the task is simply to examine an image and determine objects of interest within it. Such a tool can be used to navigate through large digital libraries to perform image analysis, cataloging of image contents, and browsing by specifying examples. In addition to automating laborious and visually-intensive tasks, it provides the means for an objective, examinable, and repeatable process for detecting and classifying objects in images.

1.3 The Query Formulation Problem

Work on techniques for digital libraries has focused mainly on digitization techniques, storage and retrieval mechanisms, and database type issues dealing with efficient indexing and query execution. We believe there is an important and crucial problem that needs to be addressed before collections of digital images can be turned into *useful digital libraries*, namely the query formulation problem. Users would like to be able to use a digital image library to search for particular patterns for cataloging or investigative purposes. A typical query would be something like: "in how many images does this pattern occur?" or "catalog all occurrences and sizes of objects in images satisfying certain conditions." Unfortunately, unlike the Case where one is dealing with a relational database or the text of a book, there is no easy way for the user to formulate the required query.

We propose an approach that calls for developing a system that learns *from examples*. Hence rather than issuing queries, the user simply provides training examples and then asks the system to "find all objects that look like this." The approach promises to bypass the query formulation bottleneck in the way humans currently interact with large databases. For most interesting tasks of image analysis, formulating queries to specify a set of target objects/regions requires solving difficult, problems that often involve effectively translating human visual intuition into pixel-level algorithmic constraints. This is a fairly challenging task in its own right. Querying a database by providing examples and counter-examples forms a novel and powerful basis for a new generation of intelligent database interface tools. These tech, capable of performing "query by content" type operations on large image databases, promise a fundamental paradigm shift of the interface between scientists and image databases, and could enable Orders of magnitude improvements in both the quantity and quality of scientific analyses of digitized image libraries.

To ground the discussion, we provide two illustrative examples of current projects at JPL, involving the development of image exploration algorithms and tools with built-in classification learning components. We first introduce the SKICAT system used to automatically generate a large comprehensive sky object catalog from a set of sky survey images at a major astronomical observatory. The second system, called JARtool, is intended for use on the Magellan collection Of over 30,000 SAR images of the surface of Venus. Both systems use machine learning techniques and are trained by the scientists to perform the analysis work. The SKICAT results demonstrate how powerful a learning system can be, resulting in performance that exceeds that of human astronomers in recognizing faint sky objects. The paper concludes with a general discussions of learning in digital image libraries and the special advantages and challenges that image databases provide.

2 CASE STUDY 1: THE SKICAT SYSTEM

The Sky Image Cataloging and Analysis Tool (SKICAT) has been developed for use on the images resulting from the 2nd Palomar Observatory Sky Survey (POSS II) conducted by the California Institute of Technology (Caltech). The photographic plates are digitized at the Space Telescope Science Institute, resulting in about 3,000 digital images of 23,040" x 23,040 pixels each, totalling over three terabytes of data. When complete, the survey will cover the entire northern sky in three colors, detecting virtually every sky object down to a B magnitude of 22 (a normalized measure of object brightness). This is at least one magnitude fainter than previous comparable photographic surveys. We estimate that, at least 5×10^7 galaxies and 2×10^9 stellar objects (including over 10^5 quasars) will be detected. This data set will be the most comprehensive large-scale imaging survey produced to date and will not be surpassed in scope until the completion of a fully digital all-sky survey.

The purpose of SKICAT is to facilitate the extraction of meaningful information from such a large database in an efficient and timely manner. The first step in analyzing the results of a sky survey is to identify, measure, and catalog the detected objects in the image into their respective classes. Once the objects have been classified, further scientific analysis can proceed. For example, the resulting catalog may be used to test models of the formation of large-scale structure in the universe, probe Galactic structure from star counts, perform automatic identifications of radio or infrared sources, and so forth.[11,36,37] Reducing the images to catalog entries is an overwhelming task which inherently requires an automated approach. The goal of this project is to automate this process, providing a consistent and uniform methodology for reducing the data sets. This will provide the means for objectively performing tasks that formerly required subjective and visually intensive manual analysis.

An important goal of this work is to classify objects whose intensity (isophotal magnitude) is too faint for recognition by inspection, hence requiring an automated classification procedure. Faint objects constitute the majority of objects on any given plate. We target the classification of objects that are at least, one magnitude fainter than objects classified in previous surveys using comparable photographic material. We shall briefly give a general, high-level description of the application domain, and report on the successful results which exceeded our initial accuracy goals. We therefore do not provide much of the details of either the learning algorithms or the technical aspects of the domain. We aim to point out an instance where learning algorithms proved to a be useful and powerful tool in the automation of scientific data analysis.

2.1 Classifying Sky Objects

SKICAT provides an integrated environment "for the construction, classification, management, and analysis of catalogs from large-scale imaging surveys. Due to the large amounts of data being collected, a manual approach to detecting and classifying sky objects in the images is infeasible: it would require on the order of tens of man years. Existing computational methods for classifying the images would preclude the identification of the majority of objects in each image since they are at levels too faint for traditional algorithms or even manual inspection/analysis approaches. A principal goal of SKICAT is to provide an effective, objective, and examinable basis for classifying sky objects at levels beyond the limits of existing technology.

Figure 1 depicts the overall architecture of the SKICAT plate catalog construction and classification process. Each plate is subdivided into a set of partially overlapping frames. Low-level image processing and object separation is performed by a modified version of the public domain FOCAS image processing software.[22,34] The image processing steps detect contiguous pixels in the image that are to be grouped as one object. Attributes are then measured based on this segmentation. The total number of attributes measured for each object by SKICAT is 40, including magnitudes, areas, sky brightness, peak values, and intensity weighted and unweighted pixel moments.

Once all attributes, including normalized and non-linear combinations of these attributes, are measured for each object, final classification is performed on the catalog. Our current goal is to classify objects into four major categories, following the original scheme in FOCAS: *star, star with fuzz, galaxy*, and *artifact*. We may later refine the classification into more classes, however, classification into one of these four classes represents adequate discrimination for primary astronomical analyses of the catalogs.

2.2 Classifying Faint Objects

In addition to the scanned photographic plates, we have access to CCD images that span several small regions in some of the frames. The main advantage of a CCD image is higher resolution and signal-to-noise ratio at fainter levels. Hence, many of the objects that are too faint to be classified by inspection on a photographic plate are easily classifiable in a CCD image. In addition to using these images for photometric calibration of the photographic plates,

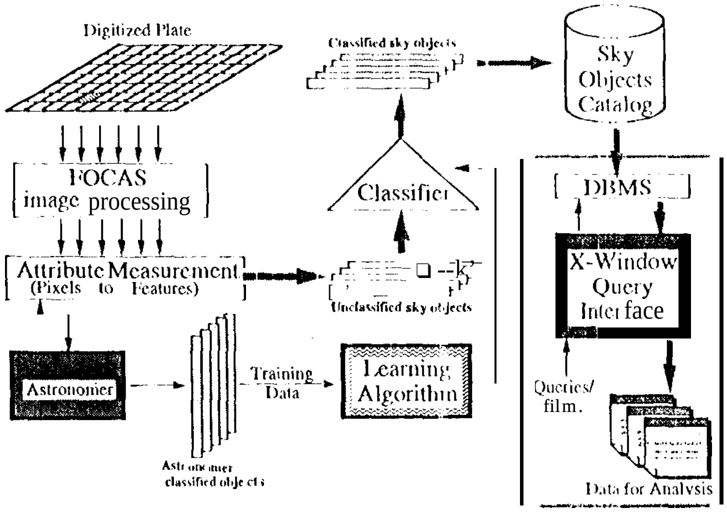

Figure 1: Architecture of the SKICAT Cataloging and Classification Process

we make use of CCD images for two machine learning purposes: 1. they enable us to obtain class labels for faint objects in the photographic plates, and 2. accurate assessment of our classifier results.

In order to produce a classifier that classifies faint objects correctly, the learning algorithm needs training data consisting of faint objects labeled with the appropriate class. This class label is obtained by examining the CCD frames. Once trained on properly labeled objects, the learning algorithm produces a classifier that is capable of properly classifying objects based *on* the values of the attributes provided by FOCAS. Hence, in principle, the classifier will be able to classify objects in the photographic image that are too faint for an astronomer to classify by inspection. We target the classification of sky objects that are at least one magnitude fainter than objects classified in photographic all-sky surveys to date.

2.3 Results and Future Work

The training and test data consisted of objects collected from four different plates from regions for which we had CCD image coverage (since this is data for which accurate classification is available). The learning algorithms are trained on a data set from 3 plates and tested on data from the remaining plate for cross validation. This estimates our accuracy in classifying objects across plates. Note that the plates cover different regions of the sky and that CCD frames cover multiple, very small portions of each plate. The training data consisted of 1,688 objects that were classified manually by an astronomer by examining the corresponding CCD frames. It is noteworthy that for the majority of these objects, the astronomer would not be able to reliably determine the classes by examining the corresponding survey (digitized photographic) images. All attributes used by the learning algorithms are derived from the survey images and not, of course, from the higher resolution CCD frames.

The learning algorithms used are based on efficiently generating decision trees or rules from the training data. It is beyond the scope of this paper to cover the algorithms. Interested readers are referred to [14,15] for a discussion of the algorithms. The classification results may be summarized as follows: Decision tree learning algorithms GID3*[16] and O-Btree[13] performed in the range of 91 % accuracy of prediction. By using RULER,[14] a program that generates many decision trees and optimizes them by extracting a robust set of classification rules via cross-validation, statistical pruning, and greedy covering, a stable result Of 94% accuracy has been achieved. For comparison, a commercially available decision tree learning algorithm called ID3 (or C4.5)[27] achieves only about 76% accuracy on average.

Note that such high classification accuracy results could only be obtained after expending significant effort on defining more robust attributes that captured sufficient invariances between various plates. When the same experiments were conducted using only the base-level at tributes measured by FOCAS, the results were significantly worse. The error rates jumped above 20% for ()- BThee, above 25% for GID3*, and above 30% for]])3. The respective sizes of the trees grew significantly as well.[14]

'J 'he SKI CAT project represents a step towards the development of an 01)jective, 1 eliable automated sky object classification method. 'J 'he initial results of our effort to automate sky object classification in order to automatically reduce the images produced by 1'OSS-11 to sky catalogs a1 e indeed vei y en couraging. Using machine learning techniques, SKICAT classifies objects that a1 e at least one magnitude fainter than objects cataloged in previous surveys. This results in a 2(K)% increase in the numbei of classi fied sky objects available for scientific analysis in the resulting sky catalog database. Furthermore, we have exceeded our initial accuracy target of 90%. This level of accuracy is required for the data to be useful in testing or refuting theoi ies on the formation of large structure in the universe and on other phenomena of interest to asti onomers. SKI CAT is now being employed to both reduce and analyze the survey images as they arrive from the digitizati on instrument. We are also beginning to explore the application of SKI CAT to the analysis of other su veys planned by NASA and other institutions.

A consequence of the SKICAT work is a fundamental change in the notion of a sky catalog from the classical static entity "in print", to a dynamic, ever growing, ever imp1 oving , on-line database. An important feature of the survey analysis system will be to facilitate suc1 i detailed interactions with the catalogs. 'J 'he catalog generated by SKICAT will eventually contain about a billion entries representii ig hundi eds of millions of sky objects. We view our effort as targeting the development of a new generation of intelligent scientific analysis tools.[37,11] Without the availability of these tools for the first survey (1'0ss-1) conducted ovei foui decades ago, 110 objective and comprehensive analysis of the data was possible. In contrast, we are targeting a comprehensive sky catalog that will be available on-line for the use of the scientific community.

As part of our plans for the future we plan to begin investigation of the app licability of unsupervised learning (clustering) techniques such as AUTOCLASS[7] to the problem of discovering clusters or groupings of interesting objects. The initial goals will be to answer questions like: 1. Are the classes of sky objects used currently by astronomers justified by the data: do they naturally a1 ise in the data? '2. Are there other classes of objects that astronomers were not aware of because of the difficulty of dealing with high dimensional spaces defined by the various attributes? The longer term goal is to evaluate the utility of unsupervised learning techniques as an aid for the types of analyses astronomers conduct after objects have been classified into known classes.

3 CASE STUDY 2: VOLCANO DETECTION IN MAGELLAN-VENUS 1)ATA

The Magellan-Venus data set constitutes yet another example of the large volumes of data that today 's instruments can collect, providing more detail of Venus than was previously available from Pioneer Venus, Venera 15/16, or ground-based radar observations put together.29 The Magellan spacecraft ti ansmitted back to earth a data set consisting of over 30,000 high resolution (75 225m pei pixel), 1024 pixel square, synthetic aperture radar (SAR) images of the Venusian surface. Planetary scientists are literally swamped by data.

We are targeting the automated detection of the "small-s hield" volcanoes (less than 15km in diameter) that constitute the most abundant visible geologic feature on the surface of Venus.[19] Identifying and studying these volcanoes is fundamental to a p roper understanding of the geologic evolution of Venus. Central to volcanic studies is the cataloging of each volcano location, its size, and characteristics. It is estimated, based on extrapolating from previous studies and knowledge of the underlying geologic p rocesses, that there should be 011 the order of 10^6 of these volcanoes visible in the Magellan data[2,21] (scattered throughout the 30,000 images). Furthermore, it has been estimated that manually locating all of these volcanoes would require 011 the order of 10 lnall-years of a planetary geologist's ti]ncto carry out,. However, locating and parameterizing them in a manual manner is forbiddingl y time-consuming. H ence, we have undertaken the development Of techniques to partially automate this task. An example image showing volcanoes appears in Section 4.'2

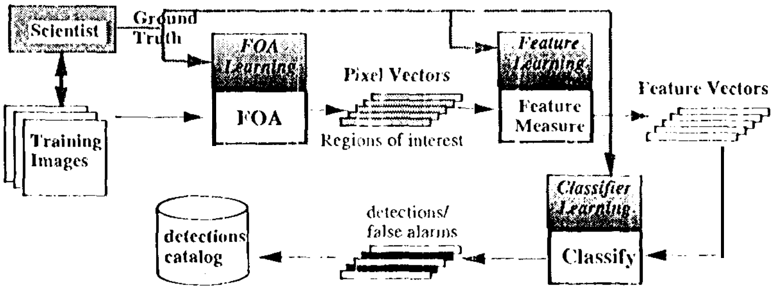

Figure 2: Block Diagram of the JARtool System

3.1 General Approach and Initial Results

There has been little prior work on detecting naturally occurring objects in remotely-sensed images. Much of the contemporary work in computer vision is model-based. While this works well for detecting *man-made* objects, the model-based approach often deals poorly with the variability in appearance of natural objects and the noise present in typical remotely sensed data.[19, 26] Hence for example, while it is impractical to specify an effective model for volcano detection based on prior knowledge, it is much more straightforward to have the scientists provide examples of volcanoes and consequently try to *learn* the mapping from pixel space to volcano/non-volcano categories.

We are developing a system called JARtool (JPL Adaptive Recognition Tool) that consists of three distinct components: focus of attention, feature learning/extraction, and classification learning. Figure 2 gives a block diagram of the approach. Our initial work in this problem has relied on the concept of using a focus of attention (FOA) method to detect regions of interest followed by local classification of regions of interest into volcano and non-volcano categories. The focus of attention component is designed primarily for computational efficiency. Its function is to quickly scan an input image and roughly determine regions of interest (regions potentially containing objects similar to those specified by the scientist). For this purpose we have used a matched filter which is automatically constructed from the training data by taking a normalized average of all volcanoes in the training set.[4] This approach detects the majority of volcanoes (including all of the volcanoes for which the scientists are most confident, in their labeling). False alarms are caused by craters, grabens, other bright features, and SAR noise.

Given a set of detected regions of interest, one must then discriminate between the volcanoes and the false alarms. A current focus of the research is to find a useful feature-representation space - a representation based purely on pixels will tend to generalize poorly. For the purposes of incorporating prior knowledge the ideal feature set would be expressed in the form of expected sizes, shapes, and relative geometry of slopes and pits, namely, the same features as used by the scientists to describe the volcanoes. However, due to the low signal-to-noise ratio of the image, it is quite difficult to accurately estimate these features, effectively precluding their use at present. The current focus of our work is on a method which automatically derives robust feature representations (see Section 4.1 for more details).

We have constructed several training sets using 75m/pixel resolution images labeled by the collaborating geologists at Brown University to get an initial estimate of the performance of the system. The FOA component typically detects more than 85% all the volcanoes the 15% which are not detected are ones which the scientists have labeled as "marginal" or lower probability volcanoes. Since it is designed to act as an aggressive filter, the FOA component generates 5 to 6 times as many false alarms as true detections. Using a maximum-likelihood Gaussian classifier, and features derived via a principal component decomposition of local pixel windows (see Section 4.1), initial cross-validation results on a particular region of the planet have shown that the system can classify volcanoes with similar accuracy to that of scientists on the same data.[4]

4 LEARNING ISSUES IN IMAGE DATA BASE ANALYSIS

Having discussed the particular details of both the POSS-II and Magellan-Venus databases and the particular systems we have implemented, the remainder of the paper will focus on the general issues which arise in problems of this nature.[17] The focus will be on some of the complicating factors which arise when image analysis and learning algorithms are combined: the role of prior information, lack of absolute ground truth, modeling spatial context, online learning and adaptation, and multi-sensor/thematic map data.

4.1 The Role of Prior Information

In general, prior information about an image exploration problem can be specified in two ways. The first is in terms of relatively high-level knowledge specifying expectations and constraints regarding certain characteristics of the objects of interest. For example, in the Magellan-Venus problem the incidence angle of the synthetic aperture radar instrument to the planet's surface is known, which in turn strongly influences the relative positions of bright and dark slope and summit regions for a given volcano.[25] The second type of prior information is the information which is implicitly specified by the labeled data, i.e., the data which have been examined by the domain expert and annotated in some manner.

One must determine the utility of each type of information in designing an image exploration algorithm. For example, in the SKICAT project, the prior knowledge was quite precise and helped a great deal in terms of determining the optimal features to use for the problem. In contrast, for the Magellan-Venus problem, the prior knowledge is quite general in nature and is not easily translatable into algorithmic constraints, leaving us only with the labeled training examples provided by the scientists.

Raw pixel data rarely provide a good basis for learning — appropriate pixel-derived *features* can typically provide a much more robust representation. In scientific data analysis, where the user typically knows the data well and has a list of defined features, selecting the appropriate feature to present to the learning algorithm can be straightforward. SKICAT provides an excellent example of this. Not only was the segmentation problem (locating objects) easy to perform, but we had access to a host of defined attributes that we made use of effectively. Having the proper representation made the difference between success and failure.

In order for SKICAT to achieve stable classification accuracy results on classifying data from different plates, we had to spend some effort defining normalized attributes that are less sensitive to plate-to-plate variation. These attributes are computed automatically from the data and are defined such that their values would be normalized within and across images and plates. Many of these quantities (although not all) have physical interpretations. Other quantities we measured involved fitting a template to a set of "sure-stars" selected by the astronomer for each image, and then measuring the rest of the objects with respect to this template. In order to automate the measurement of such attributes, we automated the "sure-star" selection problem by treating it as a learning sub-problem and building decision trees for selecting "sure-stars" in an arbitrary image. Fortunately, this turned out to be a relatively easy learning task: our accuracy on this subproblem exceeds 98%. This allowed us to automate the measurement of the needed and more sophisticated attributes. In this case a wealth of knowledge was available to us in terms of attributes (measurements), while astronomers found it *difficult to use* these attributes to classify objects. The machine learning algorithms were able to produce a classifier that used many (as many as eight) dimensions. No projection of the data onto two or three dimensions would have allowed as accurate a classification, explaining the difficulty humans found in designing the classifier.

On the other hand, in the case of the Magellan-Venus data, the feature extraction problem is significantly more difficult to address. One approach we have been experimenting with is the use of principal component analysis.[33] Each training example (a subimage containing a positive example) can be turned into a vector of pixel values. The entire training set of n examples, each of which consists of a $k \times k$ pixel subimage, will thus form a $k^2 \times n$ matrix which can subsequently be decomposed into a set of orthonormal eigenvectors using singular value decomposition (SVD). An eigenvalue is associated with each of the vectors indicating its relative importance. When the eigenvectors

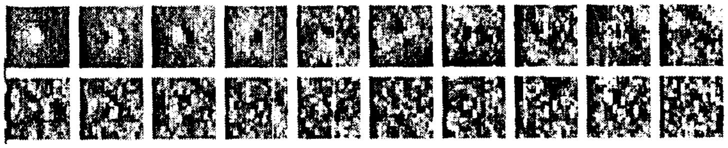

Figure 3: Eigenvolcanoes Derived from Training Data.

(eigenvolcanoes) are viewed as images again, we note that each represents a "basic" feature of a volcano. Figure 3 shows 20 associated eigenvolcano features (those corresponding to the largest eigenvectors) ordered left to right by decreasing eigenvalues. Note that the eigenvectors become less coherent starting with the sixth or seventh feature. Each block in the figure corresponds to a 225-component eigenvector that was retranslated into a 15 x 15 image and redisplayed as a block in the figure.

The eigenvolcanoes can be viewed as general features that can be used to encode each detected candidate volcano for classification purposes. This is an example of an automatic template (matched filter) generation procedure which can easily be augmented by other features provided by the expert user. However, the drawbacks of the SVD approach are worth *mentioning*: in general, the method can be sensitive to scale, rotation, and translation distortions (although these are not significant problems for the volcano problem) and it is difficult to encode prior knowledge about the domain into the model.

4.2 Lack of Absolute Ground Truth

In statistical pattern recognition and machine learning, supervised learning implicitly assumes that the "target signal" corresponds to ground truth. In image analysis, the labeling is often not in fact "ground truth", where ground truth is taken to mean that the object of interest has had its identity ascertained by a separate image-independent measurement with near-zero ambiguity. Instead, objects of interest may be labeled in a *subjective* manner by a scientist. If the signal-to-noise ratio (SNR) of the images is high enough then the subjective estimates may be accepted as near enough to ground-truth for practical purposes. This is the case with the SKICAT data, especially with the use of higher resolution CCD images to obtain training data. However for the Magellan-Venus data, with low SNR, there can be some variability between different scientists (and the same scientist at different times) in labeling a particular image. in such a case, treating such subjective estimates as ground truth is to ignore a potentially important source of noise in the data.

It is an important point that, in the absence of absolute ground truth, the goal of our work is to be as comparable in performance as possible to the scientists in terms of labeling accuracy. Absolute accuracy is not measurable for this problem. Hence, the best the algorithm can do is to emulate the scientist's performance. Also, the learning algorithms should be able to make use of probabilistic class labels during learning.[31]

Given that the scientists cannot classify each object with 100% confidence, how can we assess how well our algorithms are performing? One approach is to measure the performance of individual scientists with respect to a "consensus ground truth", where the consensus data is generated by several scientists working together discussing the merits of each candidate volcano. The performance of an algorithm is considered to be satisfactory if, compared to consensus ground truth, its performance is as good as that of an individual scientist. The philosophy here is that if a single scientist is qualified to perform the analysis, then it is sufficient if our algorithms perform comparably.

The scientists label training examples into quantized probability bins or "types", where the probability bias correspond to visually distinguishable sub-categories of volcanoes. In particular, we have used 4 types: (1) summit pits, bright-dark radar pair, and apparent topographic slope, all clearly visible, probability 0.98; (2) only 2 of the 3 criteria in category 1 are visible, probability 0.80; (3) no summit pit visible, evidence of flanks or circular outline, probability O.(K),; and (4) only a summit pit visible, probability 0.50. The probabilities correspond to the mean probability for a particular type (the probability that a volcano exists at a particular location given that it belongs to a particular type) and were elicited after considerable discussions with the planetary geologists. The use of

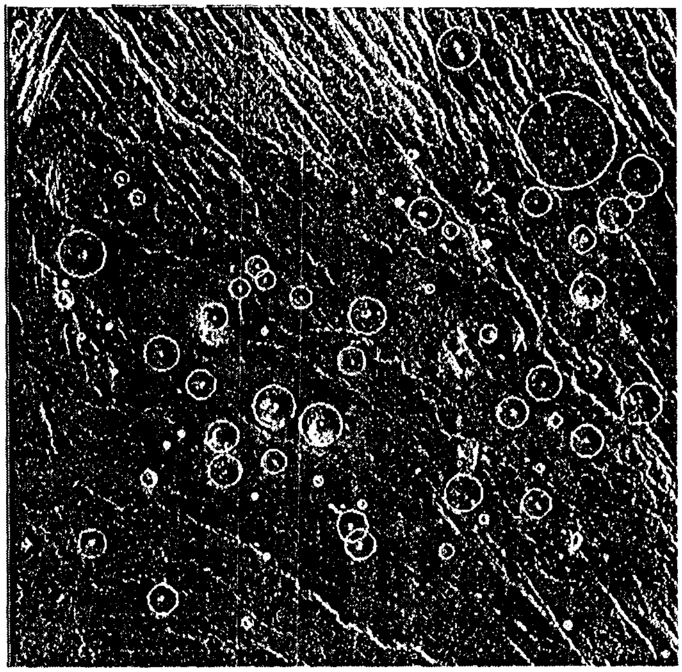

Figure 4: Magellan SAR image of Venus with consensus labels showing small volcano sizes and locations

quantized probability bins to attach levels of certainty to subjective image labeling is not new: the same approach is routinely used in the evaluation of radiographic image displays.[3,8]

We have developed a variant of the standard receiver operating characteristics (ROC) called the weighted free-response (wFROC) which takes into account both the facts that (i) the false alarm rate for detecting objects in images should be normalized relative to some quantity other than the number of pixels in the image, and (ii) that detections must be weighted both partly as detections and as false alarms, e.g., if, at a particular operating threshold, the system detects a local region which has reference probability 0.8 of being a volcano, then it must count as 0.8 of a detection and 0.2 of a false alarm.[5] More recently we have developed models which estimate the subjective labeling noise of each individual scientist, allowing the combination into a statistical estimate of the consensus of different individual labelings of the same image.[32]

Accounting for subjective uncertainty in the Magellan-Venus data set. has proven quite important. The relative performance of algorithms and scientists can change depending on whether the labels are treated as absolute ground truth or not.[5,32] Hence, modeling of subjective uncertainty can be an important component in combining image analysis and learning algorithms, particularly when dealing with low SNR images.

4.3 Online Learning and Adaptation

Another aspect of the image exploration problem is that one would ideally like to have an algorithm which could gradually improve its performance as it explores a particular database. This type of *incremental learning* has largely been ignored by researchers in learning and pattern recognition in favour of the simpler approach of "one-shot" batch learning. The particular type of model representation being used critically influences whether the model is easily adaptable or not. For example, neither discriminative models (such as the decision trees used in SKICAT) or PCA feature generation methods (as used in JARtool) are well-suited for *online* adaptation. In each case, to update the model to include new data, the training algorithm must be run in batch mode on *all* of the data which has been collected to that point. Memory and prototype-based models (including parametric density models, non-parametric density estimators, mixture models, nearest-neighbour models, etc.) are more suited to online adaptation however, they typically suffer from poor approximation properties in high dimensions.[30]

Given a particular adaptive algorithm, a unique feature of image analysis problems is the fact that the bureau visual system of the domain expert offers an excellent opportunity for supervised feedback to improve adaptation. This is in marked contrast to other domains, such as the analysis of "flat" medical data for example, where there is no intuitive way for a domain expert to visualize and quickly evaluate high-dimensional vectors for the purpose of labeling them. Hence, in principle, online image exploration algorithms could operate by iterative interaction with the human user, sequentially selecting examples from the database for labeling. While such algorithm have significant potential for changing the way in which scientists interact with image databases, it is also clear that until various fundamental representational issues are solved (cf. learning mappings from pixels to meaningful categories), such adaptive algorithms will remain beyond our reach

4.4 Modeling Spatial Context

Most learning algorithms implicitly assume that the training data consists of independent randomly chosen samples from the population of interest, e.g., a set of medical records for a hospital. Hence, such algorithms may not be directly suited to the task of learning from image data where there may be significant inter-pixel aud inter-object spatial correlation present. To handle such correlation one can impose spatial smoothness constraints on both the labels and the pixel intensities using models such as Markov *random fields* (MRF's)[18,28] and other, more global, models of spatial context.[24] In any of these approaches there is little theory on how to reconcile prior knowledge regarding spatial constraints with one's choice of model parameters, so that considerable experimentation aud tuning is often necessary for a given application.

4.5 Multi-Sensor and Derived Map Data

A common feature in remote-imaging applications is the illumination of the same target area in different ways (e.g., at multiple wavelengths), thus obtaining a vector of intensities at each pixel site rather than just a single intensity. For example, in SKICAT, the data was collected in three optical color bands. In the Magellan-Venus data, many parts of the planet were imaged from different angles and at different resolutions. In addition, low-resolution altimeter data was also measured. This results in several different data sets being available for the same surface regions. Similarly, after data has been acquired aud archived, different research groups will typically analyst the data aud produce thematic maps and catalogs (either by manual or automated means) for different quantities of interest.[6,23] For example, in the Magellan-Venus database, catalogs have already been produced for large volcanic structures and for the location of many of the large volcanic fields (but, not the volcanoes within the fields).

Hence, in the general sense, each pixel can have a vector of associated attributes, whether these are data from another sensor, or derived qualitative categories (such as a map). In principle, such additional data should be particularly useful for computer-aided detection since it is often difficult for a human user to visualize such multi-dimensional representations. However, certain technical difficulties must be overcome for the additional data to be useful. For multi-sensor data, the different data sets must usually be *registered* so that the pixel measurements are somehow aligned to reference the same surface point this can be an imprecise process. Similarly, subjectively derived thematic maps may be subject to various biases or systematic errors. Hence, methodologies for determining the relative reliability of different sources of information are highly desirable, although not always available in practice. At a minimum, automated cataloging systems and thematic mappers should provide a calibrated estimate of the reliability of the decision at each pixel or region of interest ("spatial error bars"). Although not a feature of the first-generation SKICAT or JARtool systems, probabilistic models for assessing model reliability are currently being pursued.

5 SUMMARY AND CONCLUSION

Natural object detection and characterization in large digital image libraries is a generic task which poses many challenges to current pattern recognition and machine learning methods. This paper has briefly touched on a number of relevant issues in problems of this nature. There are many other issues which impact the integration of learning and image analysis algorithms which were not discussed here due to space constraints, including the use of physical noise models for radar imaging processes and other wavelengths, the integration of multiple images of the same surface area taken at different times, and the use of multi-resolution and parallel algorithms to speed computation.

The SKICAT and JARtool projects are typical examples of the types of large-scale image database applications which will become increasingly common - for example, the NASA Earth Observing System Synthetic Aperture Radar (EOS SAR) satellite will generate on the order of 50 Gbytes of remote sensing data per hour when operational.[35] In order for scientists to be able to effectively utilize these extremely large amounts of data, basic digital image library navigation tools will be essential.

Our existing JPL projects have so far demonstrated that efficient and accurate tools for natural object detection are a realistic goal provided there is strong prior knowledge about how pixels can be turned into features and from there to class categories. With the astronomy problem there was sufficient strong knowledge for this to be the case. With the volcano data, the knowledge is much less precise and consequently the design of effective object detection and recognition tools is considerably more difficult. The common thread across the various issues would appear to be the problem of how to combine both prior knowledge and data. Much of the prior knowledge of a domain scientist is vague and imprecise and cannot be translated easily into pixel-level constraints. However, scientists find it significantly easier to provide attributes to measure on a given region than to specify the method they use to classify the region. This is an important aspect that can be exploited to solve significant problems as was done with the SKICAT system. This appears to be a good solution to the query formulation problem, which would be a major hurdle standing in the way of turning a large image database into a digital library: be it via data reduction (as in SKICAT cataloging) or object detection and recognition as in JARtool. The latter appears to be an essential problem to solve if the goal of a realistic query-by-content type capability is to be achieved.

Dealing with image data is uniquely appropriate for interactive tools since results can immediately be visualized and judged by inspection. This makes obtaining feedback and training data from users much easier. Since humans find it particularly difficult to express *how* they perform visual detection and classification, using a "learning from examples" approach becomes particularly appropriate. The fact that the image databases are becoming increasingly common and unmanageably large makes the need for the type of approaches advocated in this paper particularly pressing.

6 ACKNOWLEDGMENTS

The authors would like to thank Jayne Aubele and Larry Crumpler of the Department of Geological Sciences, Brown University, for their assistance in labeling images, and Maureen Burl (JPL) for assistance in obtaining the experimental results described in this paper. The SKICAT work is a collaboration between U. Fayyad (JPL) and N. Weir, S. Djorgovski (Caltech Astronomy). The work on JARtool is a collaboration between U. Fayyad, P. Smyth (JPL) and M. Burl, P. Perona (Caltech's E. E. Department). The work described in this paper was carried out in part by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

7 REFERENCES

[1] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *J. American Statistical Association,* 86(414), pp.376-387, June 1991.

[2] J. C. Aubele and E. N. Slyuta, "Small domes on Venus: characteristics and origins," in *Earth, Moon and Planets*, 50/51, 493-532, 1990.

[3] P. C. Bunch, J. F. Hamilton, G. K. Sanderson and A, 11. Simmons, "A Free-Response approach to the measurement and characterization of radiographic observer performance," *J. Appl. Photo. Eng.*, vol.4, no. 4, pp.166-171, 1978.

[4] M. C. Burl, U. M. Fayyad, P. Perona, P. Smyth, and M, P, Burl, "Automating the hunt for volcanoes on Venus," *Computer Vision and Pattern Recognition Conference, CVPR-94*, Los Alamitos, CA: IEEE Computer Society Press, 1,1,.302-309, 1994.

[5] M. C. Burl, U. M. Fayyad, P. Smyth, and P, Perona, "Automated analysis of radar imagery of Venus: handling lack of ground truth," *Proceedings of the IEEE Conference on Image Processing: ICIP-94*, Austin, TX, to appear, 1994.

[6] P. A. Burrough, *Principles of Geographic Information Systems for Land Resources Assessment*, Oxford: Clarenden, 1986.

[7] P. Cheeseman et al., "Bayesian Classification," *Proc. of the 7th Nat. Conf. on Artificial Intelligence AAAI-88*, pp. 607-611, Saint Paul, MN, 1988.

[8] M. S. Chesters, "Human visual perception and ROC methodology in medical imaging," *Phys. Med. Biol.*, vol.37, 110.7, pp.1433-1476, 1992.

[9] A. M. Cross, *Int. J. Remote Sensing*, 9, no.9, 1519 1528, 1987.

[10] A. P. Dawid and A. M. Skene, "Maximum likelihood estimation of observer error-rates using the EM algorithm," *Applied Statistics*, vol.28, no.1, pp.20 28, 1979.

[11] S. G. Djorgovski, N. Weir, and U. M. Fayyad, "Processing and Analysis of the Palomar STScI Digital Sky Survey Using a Novel Software Technology", in D. Crabtree, R. Barnes, and R Hanisch (eds.), *Astronomical Data Analysis Software and Systems III, A .S.P. Conf. Ser.* in press.

[12] Duda, R.O and Hart, P.E. (1 973) *Pattern Classification and Scene Analysis.* New York: John Wiley and Sons.

[13] U. M. Fayyad and K. B. Irani, "The attribute selection problem in decision tree generation" *Proc. of the Tenth National Conference on Artificial Intelligence AAAI-92* (pp.104-1 10). Cambridge, MA: MIT Press, 1992.

[14] U. Fayyad, N. Weir, and S.G. Djorgovski, "SKICAT: a machine learning system for automated cataloging of large scale sky surveys." *Proc. of Tenth Int. Conf. on Machine Learning*, Morgan Kaufman, 1993.

[15] U. M. Fayyad and K. B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," *Proc. of the 13th International Joint Conference on Artificial Intelligence IJCAI-93.* Chambery, France: Morgan Kaufman, 1993.

[16] U. M. Fayyad, "Branching on attribute values in decision tree generation," *Proc. of the Twelfth National Conference on Artificial Intelligence AAAI-94*, pp.601-606, Seattle, WA: AAAI/MIT Press, 1994.

[17] U. M. Fayyad, P. Smyth, N. Weir, and S. Djorgovski, "Automated analysis and exploration of large image databases: results, progress, and challenges," *Journal of Intelligent Information Systems*, 4, 1-19, in press, 1995.

[18] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Int.*, vol.6, no.6, 721-741, 1984.

[19] J. E. Guest et al., *Journal Geophys. Res.*, 97, E10, 15949, 1992.

[20] J. W. Head et al., " Venus volcanic centers and their environmental settings: recent data from Magellan," EOS 72, p.175, American Geophysical Union Spring meeting abstracts, 1991.

[21] J. W. Head et al., *Journal Geophysical Res.*, ,97, E8, 13,153-13,197, 1992.

[22] J. Jarvis and A. Tyson, *Astr. Journ.* 86:41, 1981.

[23] D. F. Marble and D. J. Peuquet, "Geographical information systems and remote sensing," in *Manual of Remote Sensing*, 2nd ed., R. E. Colwell (ed.), Falls Church, VA: Amer. Sot. Photogrammetry, 1983.

[24] M. I. Miller, G. E. Christensen, and Y. Amit, "A mathematical textbook of deformable neuro-anatomics,' submitted to *Science*, 1993.

[25] G. J. Pettengill et al., "Magellan: radar performance and products," *Science*, vol.252, 260 265, 12 April 1991.

[26] S. Quegan, et al, *Trans. R. Sot. London*, A 324, 409-4?1, 1988.

[27] J. R. Quinlan, "The induction of decision trees." *Machine Learning* vol. 1, no. 1, 1988.

[28] B. D. Ripley, *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge, 1988.

[29] *Science*, special *issue on Magellan data,* April 12, 1991.

[30] P. Smyth and J. Mellstrom, "Detecting novel classes with applications to fault diagnosis," in *Proceedings of the Ninth International Conference on Machine Learning*, Morgan Kaufmann Publishers: Los Altos, CA, pp.416-425, 1992.

[31] P. Smyth "Learning with probabilistic supervision," in *Computational Learning Theory and Natural Learning Systems 3*, T. Petsche, M. Kearns, S. Hanson, R. Rivest (eds.), Cambridge, MA: MIT Press, in press.

[32] P. Smyth, U. Fayyad, M. Burl, P. Baldi, P. Perona, "Inferring ground truth from subjective labelling of radar images of Venus," to appear in *Advances in Neural Information Processing* 7, G. Tesauro, D. Touretzky, J. Alspector (eds.), Palo Alto, CA: Morgan Kaufman Publishers, to appear, 1995.

[33] M. Turk and A. Pentland, "Eigenfaces for recognition." *J. of Cognitive Neurosci.*, 3:71-86, 1991.

[34] Valdes (1982). *Instrumentation in Astronomy IV*, SPIE vol. 331, no. 465.

[35] J. Way and E. A. Smith, "The evolution of synthetic aperture radar systems and their progression to the EOS SAR," *IEEE Trans. on Geoscience and Remote Sensing*, vol.29, 110.6, pp.962-985, 1991.

[36] N. Weir, S. G. Djorgovski, U.M. Fayyad, et al., "SKICAT: A system for the scientific analysis of the Palomar-STScI Digital Sky Survey," *Proc. of Astronomy from Large databases II*, p. 509, Munich, Germany: European Southern Observatory, 1992.

[37] N. Weir, et al., "Cataloging the Northern Sky Using a New Generation of Software Technology," in H. MacGillivray (ed.), *Astronomy From Wide-Field Imaging*, Proceedings of the IAU Symp. #161, Dordrecht: Kluwer, in press.